Supporting Information

Oosterhof and Todorov 10.1073/pnas.0805664105

SI Text

Behavioral Trait Rating Studies. The faces were presented three times to increase the interrater agreement and, consequently, the reliability of judgments by reducing the measurement error for each participant. For each participant, we computed the average judgment for each of the 66 faces. The reliability of the trait judgments was computed on the standardized average ratings of participants. The sample sizes varied across studies because we wanted to achieve a criterion reliability greater than 0.90 for each trait dimension. Correspondingly, for trait judgments showing lower inter-rater agreement (e.g., "intelligent", Table S2), we collected data from more participants. For two of the traits, boring and egoistic, the interrater agreement was low (Table S2), and these traits were not included in the subsequent analyses. The final 13 mean judgments averaged across the standardized participants' ratings for each trait were submitted to a principal components analysis (PCA) to identify the underlying dimensions of face evaluation (Table S3).

The reason we standardized the mean ratings of each participant was that this procedure increases the reliability of the mean ratings across participants. To check whether this procedure affects the PCA solution (Table S3), we conducted another PCA on the mean unstandardized ratings across participants. The solution was practically identical. The correlation between the first PC derived from the analysis of the mean standardized ratings and the first PC derived from the analysis of the mean ratings was 0.999. The corresponding correlation for the second PC was 0.998.

Additional PCAs on Subsets of Trait Judgments. To test whether the frequency of use of trait terms in free, unconstrained descriptions of faces affects the PCA solution, we conducted two additional analyses. First, we weighted the PCs scores by the frequency of use of traits. Second, we conducted a series of PCAs entering at each step only a subset of the traits according to their frequency of use (Table S4). In both analyses, the correlation between trustworthiness judgments and the first PC was much less affected by these procedures than the correlation between dominance judgments and the second PC.

One reason that the first PC is more robust with respect to the sets of traits used in the PCA is that this component reflects the evaluative meaning of judgments and social judgments are inherently evaluative (1, 2). However, it is also possible that the lower correlations between dominance and the second PC can be partially explained by person positivity biases (3). People are less likely to describe other people negatively than positively. This bias would produce more positive than negative trait descriptions (Table S1). Because some of the negative traits are important (aggressive and mean, Table S3) for the dominance component, this would underestimate the stability of the correlation between dominance judgments and the second PC.

Comparing Correlations for Natural and Computer-Generated Faces. Using the same procedures as in the trait rating studies (2.1–2.15), we collected judgments of threat ($n = 18$, $\alpha = 0.91$) for the set of 66 natural faces used in the initial PCA.

For the set of 300 computer-generated faces, in addition to judgments of trustworthiness, dominance, and threat (studies 3, 4, and 13), we collected judgments of attractiveness ($n = 35$, $\alpha = 0.91$), meanness ($n = 27$, $\alpha = 0.83$), how frightening the person is ($n = 28$, $\alpha = 0.84$), extroversion ($n = 28$, $\alpha = 0.83$), likeability ($n = 32$, $\alpha = 0.75$), and competence ($n = 44$, $\alpha = 0.83$).

The patterns of correlations between the trait judgments that were made for both sets of faces (Table S5) were qualitatively the same. The signs of all correlations were identical and the magnitudes of the correlations were comparable. We also conducted a PCA on the trait judgments of the computer generated faces (Table S6). Although the set of traits was different from the set of traits used in the PCA of the judgments of the natural faces, the solution was very similar. The first PC could be interpreted as valence with trustworthiness showing the highest correlation with this component, and the second PC could be interpreted as dominance with dominance showing the highest correlation with this component.

Statistical Model for Face Representation. We used Facegen Modeller 3.1 (http://facegen.com), in which the shape of a 3D face is represented by the vertex positions (points in 3D Euclidian space) of a polygonal model of fixed mesh topology (Fig. S4). The geometry of the average shape is represented by an average shape vector $\bar{S} = (\bar{x}_1 \bar{y}_1, \bar{z}_1, \ldots, \bar{x}_n, \bar{y}_n, \bar{z}_n) \in \Re^{3n}$ for n vertices. Each of the 50 principal components is related to an offset vector

$$\Delta_i = (\Delta x_{1i}, \Delta y_{1i} \Delta z_{1i}, \ldots \Delta x_{ni} \Delta y_{ni} \Delta z_{ni}) \in \Re^{3n}$$

that represents how much each vertex must be displaced if the ith component is changed 1 SD. A face shape is represented by a shape vector $\alpha \in \Re^{50}$ that maps the component's offset vectors Δ_* to the geometry of the face S according to

$$S = \bar{S} + \sum_{i=1}^{50} \alpha_i \cdot \Delta_i.$$

Each of the shape dimensions α_* is independent and normally distributed N (0, 1).

Constructing Models of How Faces Vary on Social Dimensions. The trustworthiness dimension was based on the best linear fit of the mean judgments (study 3) as a function of the 50 shape components. We used simple linear model that is data-driven and prevents overfitting of the data (50 parameter estimates based on 300 data points). The linear model also assumes that there are no interactions between the shape dimensions. For practical purposes, going beyond this is hardly possible in a data-driven approach. For example, 50 shape dimensions with first order interactions would require estimating the values for more than two thousand regression coefficients. Modeling all interactions would require about 10^{15} coefficient estimates.

Using the trustworthiness judgments column vector $r \in \Re^{300}$ with values centered around zero and the face matrix $F \in \Re^{50 \times 300}$ (each column is a face shape vector) that contains the 300 randomly created faces, it can be shown that the optimal direction for the trustworthiness vector is given by the vector $t = F \cdot r$ (4). After normalizing this vector to

$$\hat{t} = \frac{t}{||t||}$$

for any arbitrary face with shape vector α the features specific for trustworthiness can be changed δ SD by changing the face shape vector to $\alpha' = \alpha + \delta \cdot \hat{t}$. The same logic applies to the dominance vector (u) (study 4).

Using these trustworthiness (t) and dominance (u) vectors, we constructed a new vector v that is closest to the dominance vector

u but orthogonal to the trustworthiness vector t. This new vector v lies in the plane that is spanned by the vectors t and u, or more formally, an element of the 2-dimensional subspace in $\Re^{50}$ spanned by t and u. Hence, it must hold that $v = a \cdot t + b \cdot u$ for certain values of a and b. It can be shown that $v = a \cdot (t - \langle t,u \rangle \cdot u)$. Now a can be chosen so that $\|v\| = 1$, yielding two possible solutions for v in opposite direction. The solution for v is chosen so that $\langle u,v \rangle \geq 0$ is positive, minimizing the angle between v and u. In our case, we found $a = 0.48$ and $b = 0.88$.

Using the same procedures as those for creating the trustworthiness vector, we created a threat vector based on the threat judgments of the 300 faces (study 13). We also constructed a new vector in the plane spanned by the trustworthiness and orthogonal dominance vectors t and v (note that the original, not-orthogonal dominance vector u also lies in this plane). Along the same lines as explained above, all vectors in this plane are described by weights a and b to yield a weighted sum $a \cdot t + b \cdot v$. Such vectors can be used to predict other vectors obtained from behavioral judgments. The vector that combines facial features of low trustworthiness with those of high dominance is at the midline "between" t and v (Fig. S9), with $a = -1$ and $b = 1$ and thus described by the vector $-t + v$, which was normalized to be used as a control in Facegen.

Addressing Potential Biases in the Selection of Faces. Although the faces used in studies 3, 4, and 13 were generated to be neutral, it is possible that some of them had expressions that biased the model. However, this is unlikely because the 300 faces were normally distributed around the center of the trustworthiness dimension (Fig. S7). Seventy percent of the faces were within 1 SD of the center of the dimension, 97% were within 2 SD, and 100% were within 3 SD. Extrapolating from the findings of the emotion categorization study (study 7), the 300 faces are within the range in which faces are categorized as neutral. Further, we obtained the same model solution using another set of 200 faces (*SI Text* Studies, Fig. S7).

Supporting Studies. In three supporting studies, we replicated the basic findings for the trustworthiness dimension using a set of 200 randomly generated faces (Fig. S6). We asked 26 participants to judge these faces on trustworthiness (*SI Text* Study 1). We used the mean judgments to find a vector in the 50-dimensional face space whose direction is optimal in changing trustworthiness (Fig. S6*A*).

To validate that the model successfully manipulates face trustworthiness, we asked 27 participants to judge faces generated by the model on trustworthiness (*SI Text* Study 2). Trustworthiness judgments of faces tracked the trustworthiness predicted by the model (Fig. S6*B*), $F_{\text{linear}}(1,26) = 38.23$, $P < 0.001$, although people were more sensitive to changes in trustworthiness at the low end of the spectrum than at the high end, $F_{\text{quadratic}}\ (1, 26) = 46.03$, $P < 0.001$.

To test whether faces with extremely exaggerated features on the trustworthiness dimension are perceived as expressing emotions, we asked 21 participants to classify faces with exaggerated features as neutral or as expressing one of the six basic emotions (*SI Text*, Study 3). As in study 7, the only responses above the chance level were for the categories of neutral, angry, and happy (Fig. S6*C*). The faces at the center of the dimension (0 SD) were classified as neutral. As the facial features become more exaggerated, the neutral categorization approached chance, $F_{\text{quadratic}}\ (1, 20) = 121.92$, $P < 0.001$. This was particularly clear on the negative end of the continuum (Fig. S6*D*). As the facial features become exaggerated in the negative direction, the faces were classified as angry, $F_{\text{linear}}(1,20) = 57.55$, $P < 0.001$, and as the features become exaggerated in the positive direction, the faces were classified as happy, $F_{\text{linear}}(1,20) = 73.35$, $P < 0.001$.

Study 1: Judgments of Computer-Generated Faces. ***Participants.*** Twenty-six undergraduate students participated in the study for partial course credit.

Face stimuli. We generated 200 random Caucasian faces using the procedures described in studies 3 and 4. The only difference was that in this study, we introduced a bias toward male faces. This bias in the sample of faces may explain the lower trustworthiness ratings in this study ($M = 4.07$, SD $= 0.51$) than in study 3 ($M = 4.75$, SD $= 0.66$).

Procedures. Participants were told that the faces might be shown multiple times, and that there is no right or wrong answer. The faces were presented two times in two separate blocks. Within each block, the order of faces was randomized for each participant. Each face was presented at the center of the screen for 500 ms. The interstimulus interval was 1,000 ms, and a fixation cross shown for 1,000 ms preceded each face. The response scale ranged from 1 (Not at all trustworthy) to 9 (Extremely trustworthy).

The mean judgments averaged across participants were used to find a dimension of trustworthiness in the 50-dimensional face space. The procedures for creating this dimension were the same as the procedures used to create the trustworthiness dimension based on the judgments of the 300 randomly generated faces. The trustworthiness judgments were reliable, $\alpha = 0.84$.

Study 2: Model Validation. ***Participants.*** Twenty-six graduate and undergraduate students and university staff members participated in the study for payment of $4.

Face stimuli. We generated 20 random Caucasian faces using the same procedure as in studies 5 and 6. Using the trustworthiness model, for each face, we created three untrustworthy (−4.5, −3, and −1.5 SD) and three trustworthy versions (1.5, 3.0, and 4.5 SD). This resulted in 140 faces, including the 0 SD faces.

Procedures. The procedures were the same as in study 5.

Study 3: Emotion Categorization. ***Participants.*** Twenty-one adults were recruited in a shopping mall. They were paid $5 for their participation.

Face stimuli. We generated 20 random Caucasian faces using the same procedures as in studies 5 and 6. These faces were divided into two sets of 10 faces, and the sets were used to create two versions of the experiment. Using the trustworthiness model, we exaggerated the features of each face to the extremes of the trustworthiness dimension. For each face, we created two untrustworthy and two trustworthy versions by moving the face 4 and 8 SD in both directions of trustworthiness. This resulted in 100 faces (50 faces in each experimental version).

Procedures. Participants were randomly assigned to one of the experimental versions. The procedures were the same as in study 7. The total number of trials was 50.

1. Rosenberg S, Nelson C, Vivekananthan PS (1968) A multidimensional approach to the structure of personality impressions. *J Pers Soc Psychol* 9:283–294.
2. Kim MP, Rosenberg S (1980) Comparison of two structural models of implicit personality theory. *J Pers Soc Psychol* 38:375–389.
3. Sears DO (1983) The person-positivity bias. *J Pers Soc Psychol* 44:233–250.
4. Blanz V, Vetter T (1999) in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (Addison Wesley Longman, Los Angeles), pp 187–194.

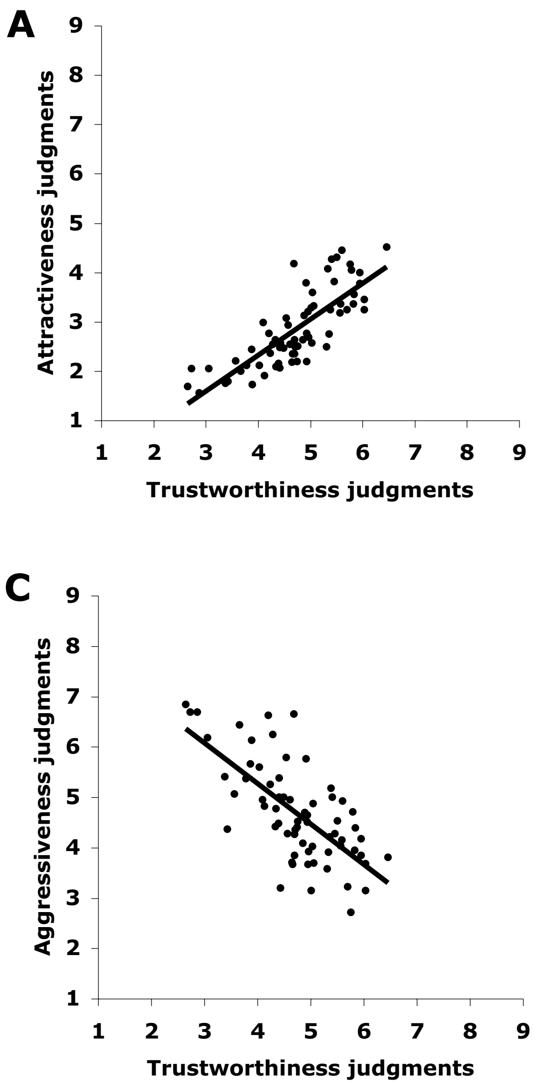


Fig. S1. Scatter plots of mean judgments of trustworthiness (n = 20) from emotionally neutral faces and (*A*) mean judgments of attractiveness (n = 21); (*B*) mean judgments of intelligence (n = 27); (*C*) mean judgments of aggressiveness (n = 19); and (*D*) mean judgments of emotional stability (n = 21). Each point is a face. Judgments were made on a 9-point scale. The lines represent the best linear fit.

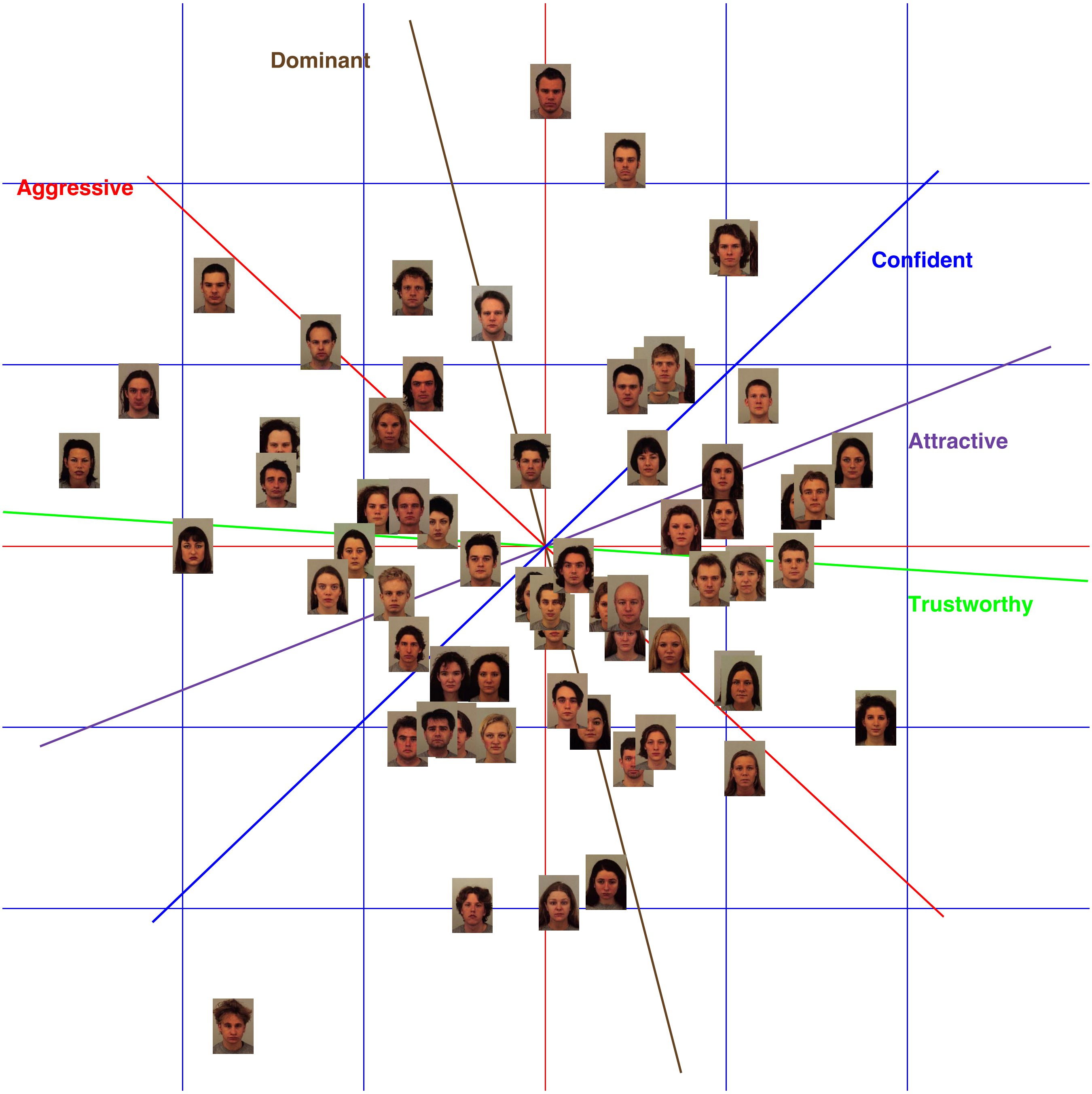


Fig. S2. Face plot of emotionally neutral faces derived from a principal component analysis of 13 trait judgments. The faces are plotted as a function of their values on the first two principal components, evaluation (*x* axis) and dominance (*y* axis). The color lines show the positions of judgments of trustworthiness, attractiveness, confidence, dominance, and aggressiveness in the two-dimensional space. The blue grid lines are spaced at standard deviation units.

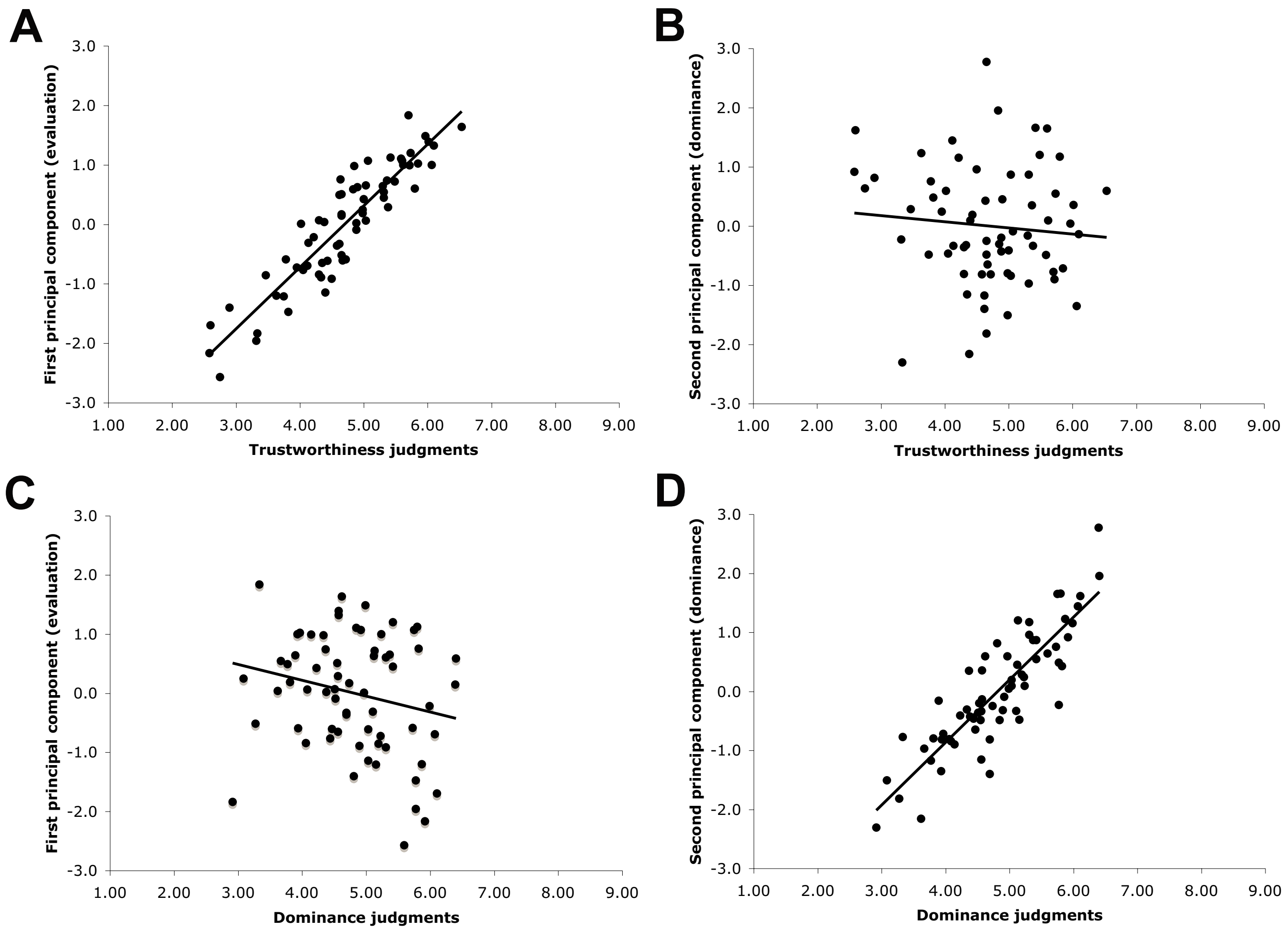


Fig. S3. Scatter plots of trustworthiness and dominance judgments of emotionally neutral faces and the first two principal components derived from a PCA of judgments on 11 traits (other than trustworthiness and dominance) used to spontaneously characterize faces (studies 1 to 2.15). (*A*) Trustworthiness judgments and the first evaluation component. (*B*) Trustworthiness judgments and the second dominance component. (*C*) Dominance judgments and the first evaluation component. (*D*) Dominance judgments and the second dominance component. Each point is a face. The judgments were measured on a 9-point scale, ranging from 1 (not at all [trustworthy or dominant]) to 9 (extremely [trustworthy or dominant]). The lines represent the best linear fit.

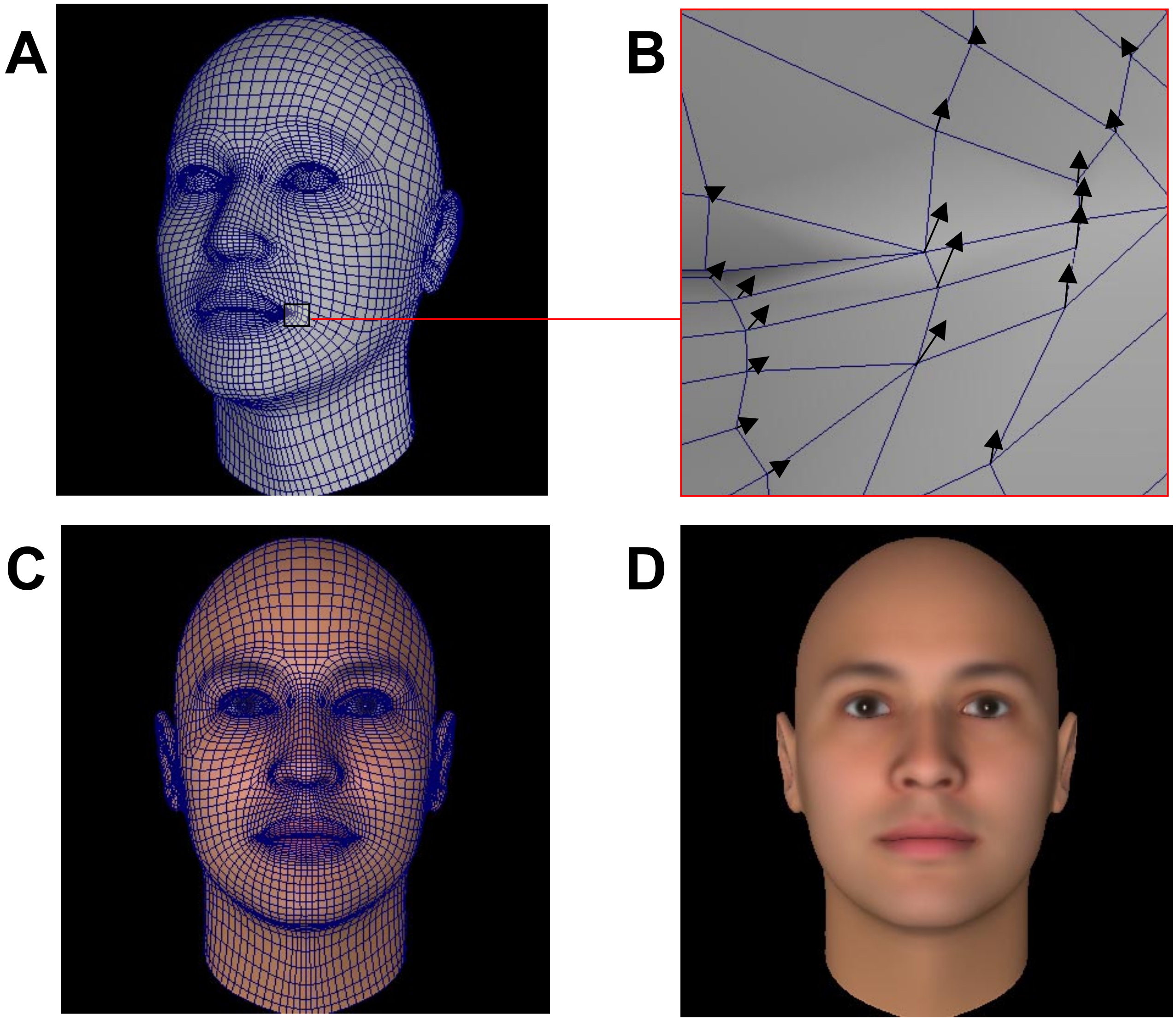


Fig. S4. Illustration of how the face model represents faces. (*A*) Side view showing the surface mesh superimposed on the average face. (*B*) Linear changes in the vertex positions for the surface of a face segment on one of the 50 shape dimensions. (*C*) Frontal view of the average face showing the surface mesh and texture. (*D*) Frontal view of the average face with texture. Face shape for specific faces is represented as a function of the average face and the differences from this face for all vertices. These differences are captured in the principal components for shape.

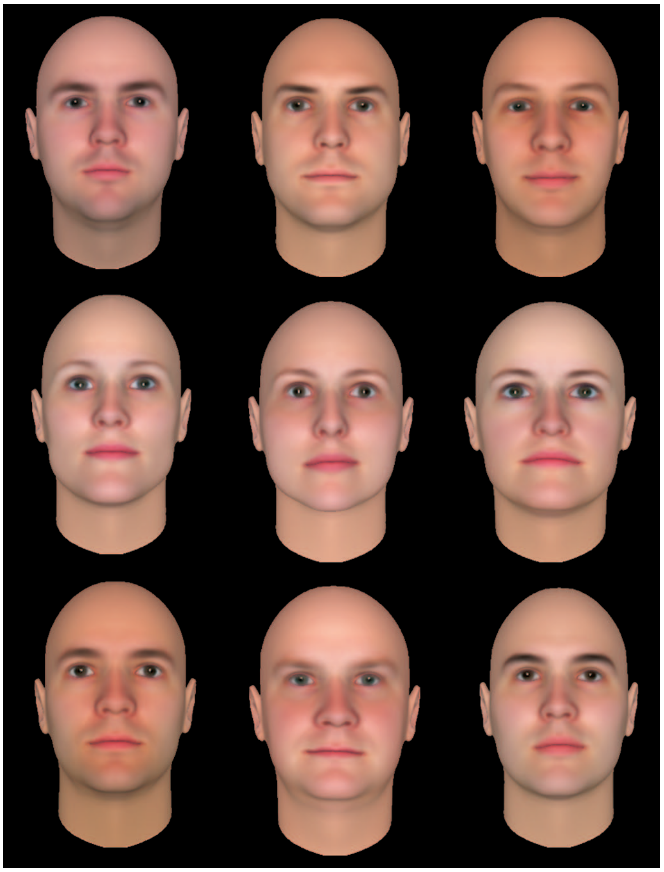


Fig. S5. Examples of computer-generated emotionally neutral faces used in studies 3, 4, and 13. The mean trait judgments of the faces were used to create vectors in the 50-dimensional face space whose direction was optimal in changing trustworthiness, dominance, and threat.

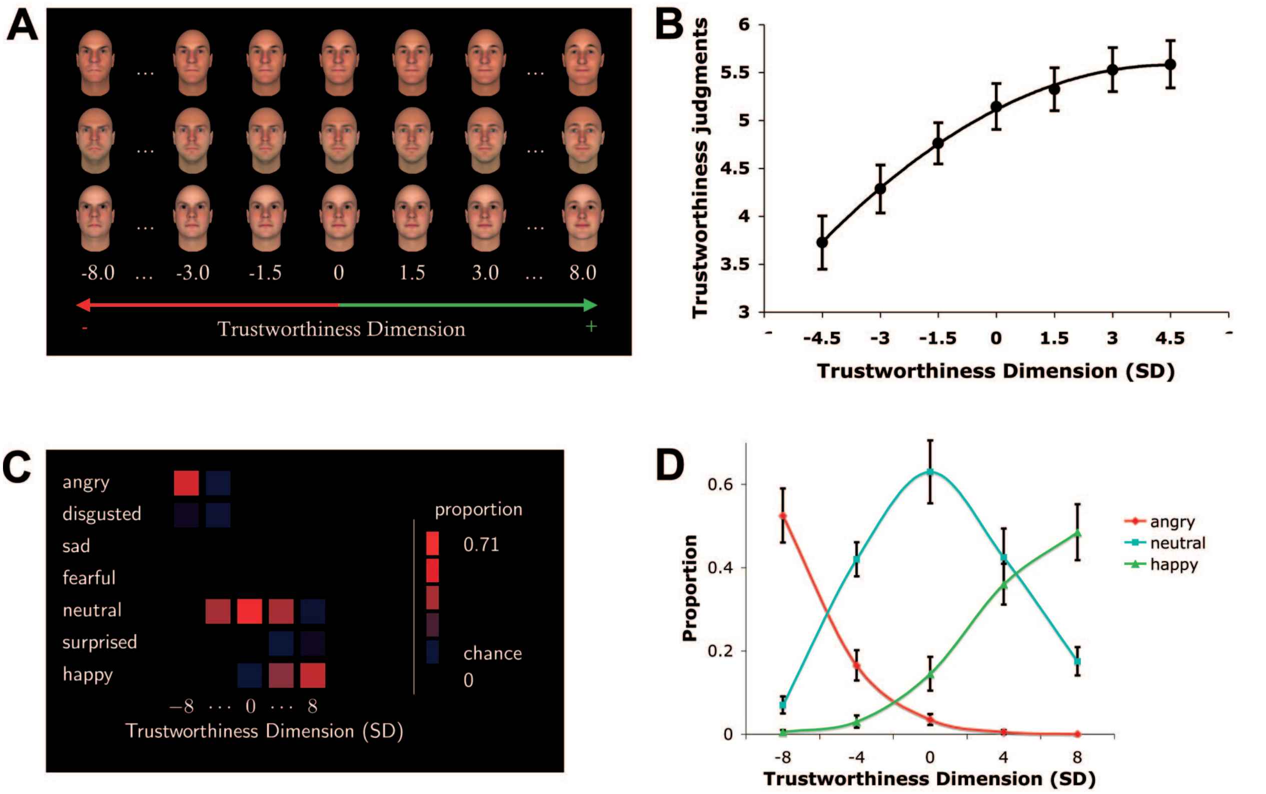


Fig. S6. (*A*) Examples of faces with exaggerated trustworthiness features. The faces in the center column were randomly generated and then their features were exaggerated to decrease (left three columns) and increase (right three columns) their perceived trustworthiness. These changes were implemented in a computer model based on trustworthiness judgments of 200 emotionally neutral faces (*SI Text* Study 1, $n = 26$). (*B*) Mean trustworthiness judgments (*SI Text* Study 2, $n = 19$) of faces generated by the computer model. The judgments were made on a 9-point scale, ranging from 1 (not at all trustworthy) to 9 (extremely trustworthy). (*C*) Intensity color plot showing the categorization of faces as neutral or as expressing one of the six basic emotions as a function of their trustworthiness (*SI Text* Study 3, $n = 21$). (*D*) Categorization of faces as angry, happy, and neutral as a function of their trustworthiness. The *x*-axis represents the extent of exaggeration of facial features in standard deviation units. Error bars show standard error of the mean.

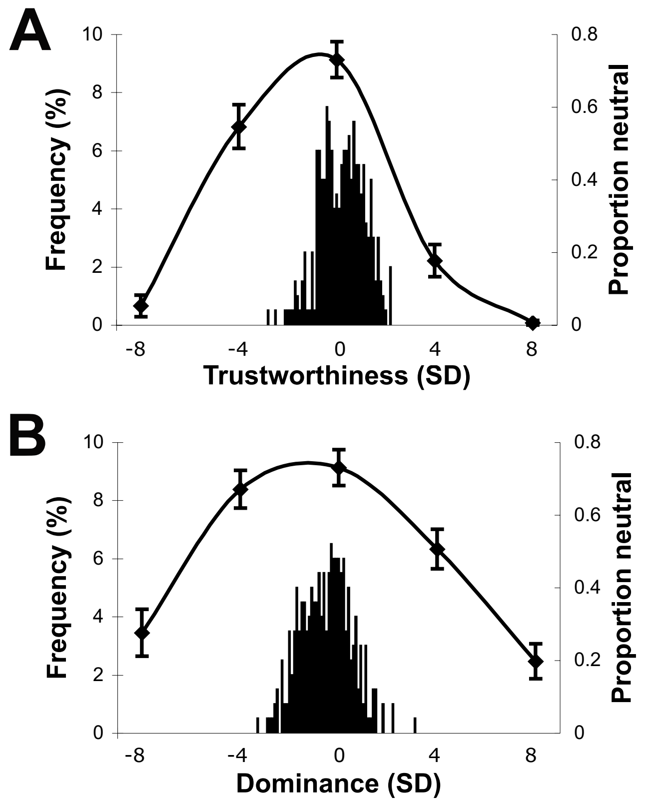


Fig. S7. Frequency distribution of the 300 faces used to create the trustworthiness and dominance dimensions (studies 3 and 4). (*A*) The frequency is plotted as a function of the values of the faces on the trustworthiness dimension. The ∩-shaped curve superimposed on the frequency distribution shows the proportion of faces categorized as neutral (on the right *y*-axis) as a function of their trustworthiness in study 7 ($n = 19$). Extrapolating from these findings, the 300 faces are within the range in which faces are categorized as neutral. (*B*) The frequency is plotted as a function of the values of the faces on the dominance dimension. The ∩-shaped curve superimposed on the frequency distribution shows the proportion of faces categorized as neutral (on the right *y*-axis) as a function of their dominance in study 7 ($n = 19$). The chance level categorization is .14. Error bars show standard error of the mean.

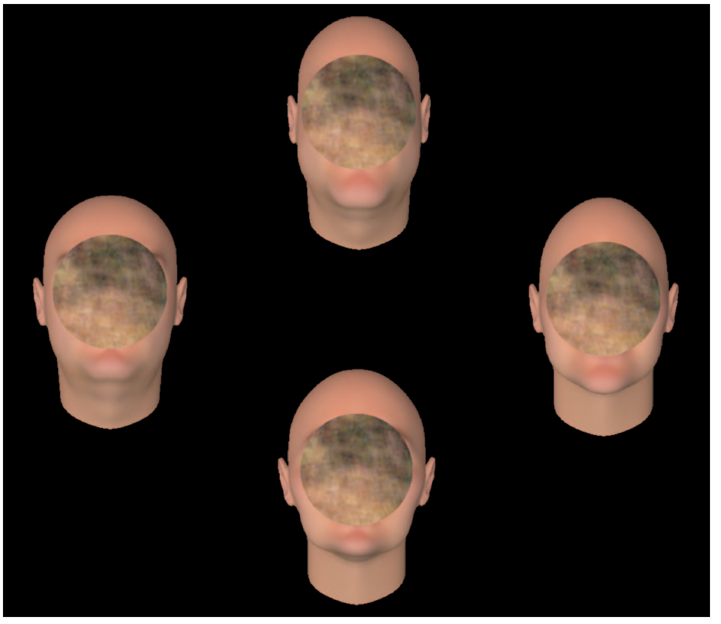


Fig. S8. Examples of face stimuli used in study 10 (n = 16). The faces on the horizontal axis represent the extremes of the trustworthiness dimension (−8 SD for the face on the left and 8 SD for the face on the right. The faces on the vertical axis represent the extremes of the dominance dimension (−8 SD for the lower face and 8 SD for the upper face).

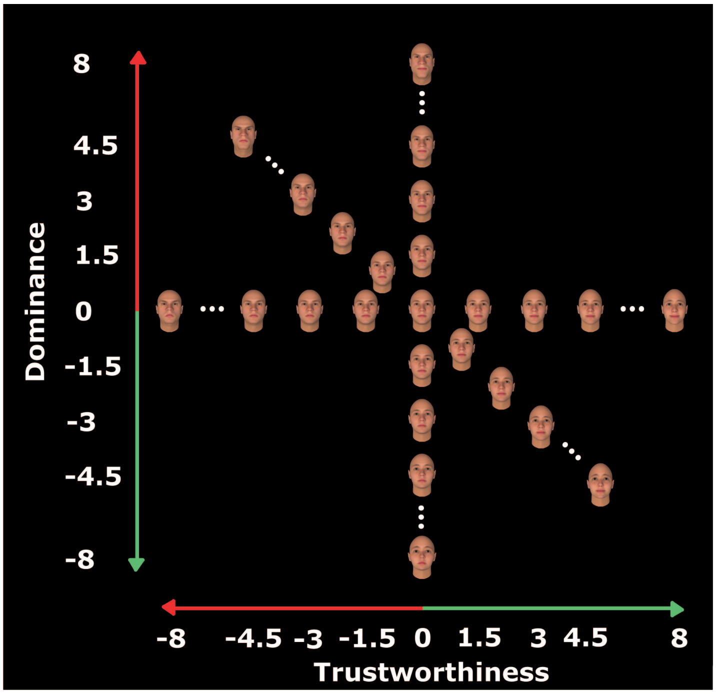


Fig. S9. Examples of a face with exaggerated features on the dimensions of trustworthiness, dominance, and threat. The threat dimension shown on the diagonal was obtained by rotating the trustworthiness dimension 45° clockwise and the dominance dimension 45° counterclockwise in the plane defined by the two dimensions. High threat faces are presented in the fourth quadrant (low trustworthiness and high dominance). Low threat faces are presented in the second quadrant (high trustworthiness and low dominance). This threat dimension was practically identical to a dimension based on threat judgments of faces (study 13, $n = 21$): one SD change on the former dimension corresponded to 0.98 SD change on the latter dimension. The extent of exaggeration of facial features is presented in standard deviation units.

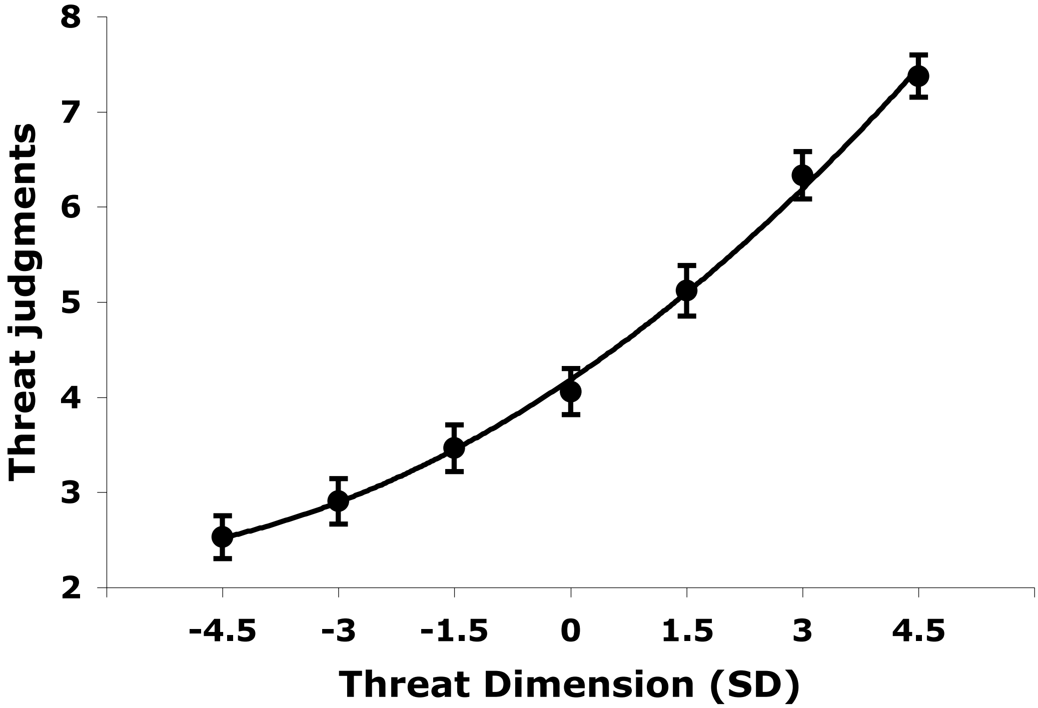
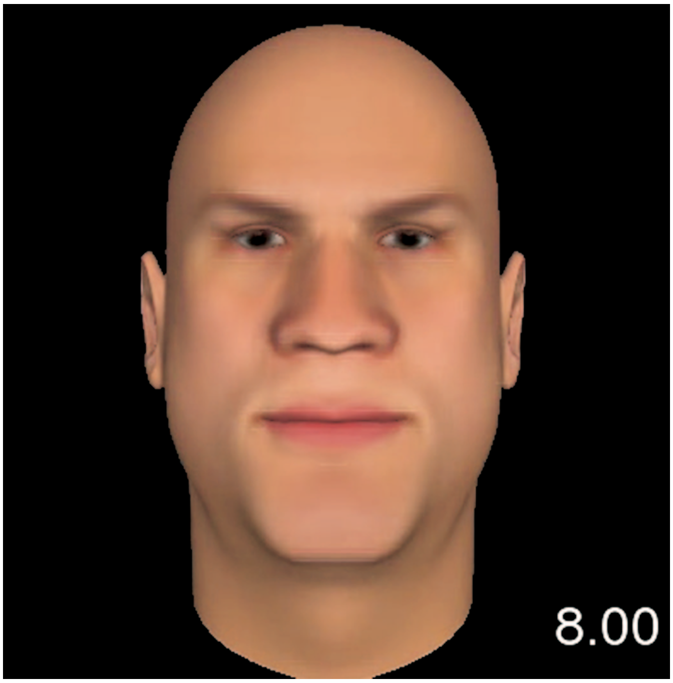


Fig. S10. Mean threat judgments of faces (study 14, $n = 18$) generated by the computer model of face threat (Fig. S9). The judgments were made on a 9-point scale, ranging from 1 (not at all threatening) to 9 (extremely threatening). Error bars show standard error of the mean.

Movie S1. Changes of facial features on the trustworthiness dimension in the range used in the behavioral studies. The running trustworthiness score is presented in standard deviation units in the lower right corner of the movie.

Movie S1(MP4)



Movie S2. Changes of facial features on the dominance dimension in the range used in the behavioral studies. The running dominance score is presented in standard deviation units in the lower right corner of the movie.

Movie S2(MP4)

Table S1. Frequency of use of trait dimensions in unconstrained person descriptions of emotionally neutral faces, and the respective means (standard deviations) of trait judgments from rating studies

Trait dimension	Frequency count	Relative proportion	Means (SD)
1. Attractive	150	0.20	2.85 (0.78)
2. Unhappy	95	0.13	4.72 (0.82)
3. Sociable	89	0.12	4.58 (0.74)
4. Emotionally stable	85	0.11	4.74 (0.79)
5. Mean	63	0.08	4.94 (0.87)
6. Boring	60	0.08	5.66 (0.49)
7. Aggressive	43	0.06	4.68 (0.98)
8. Weird	43	0.06	5.01 (1.05)
9. Intelligent	35	0.05	4.88 (0.68)
10. Confident	24	0.03	4.81 (0.68)
11. Caring	21	0.03	4.54 (0.72)
12. Egotistic	21	0.03	4.84 (0.58)
13. Responsible	20	0.03	4.31 (0.77)
14. Trustworthy	12	0.02	4.74 (0.85)
15. Dominant	–	–	4.81 (0.81)

In the rating studies, the trait judgments were measured on 9-point scales ranging from 1 (not at all [trait]) to 9 (extremely [trait]).

Table S2. Interrater agreement (r) and reliability (Cronbach's α) for judgments of emotionally neutral faces on 15 trait dimensions

Trait dimension	Interrater agreement (r)	Reliability (α)	Sample size (*n*)
1. Attractive	0.47	0.95	21
2. Weird	0.45	0.95	22
3. Mean	0.41	0.93	19
4. Trustworthy	0.39	0.93	20
5. Aggressive	0.38	0.92	19
6. Caring	0.37	0.93	22
7. Emotionally stable	0.37	0.92	21
8. Unhappy	0.36	0.93	23
9. Responsible	0.36	0.91	18
10. Sociable	0.36	0.92	20
11. Dominant	0.31	0.93	28
12. Confident	0.30	0.91	23
13. Intelligent	0.26	0.90	27
14. Egoistic	0.16	0.81	22
15. Boring	0.09	0.67	22

"Egoistic" and "boring" were not included in subsequent analyses, because the interrater agreement was low.

Table S3. Loadings of trait judgments of emotionally neutral faces on the first two principal components

Trait dimension	Valence evaluation	Dominance evaluation
Trustworthy	0.94	−0.06
Emotionally stable	0.93	0.19
Responsible	0.91	0.11
Sociable	0.91	0.20
Caring	0.90	−0.29
Weird	-.87	−0.22
Attractive	0.81	0.32
Mean	-.76	0.55
Intelligent	0.72	0.13
Aggressive	-.71	0.66
Unhappy	-.71	0.01
Confident	0.68	0.65
Dominant	-.24	0.93
Explained variance	63.3%	18.3%

The loadings represent the correlations of the trait judgments with the principal components.

Table S4. Correlations of judgments of trustworthiness and dominance with the first two principal components (PC) derived from principal components analyses of sets of trait judgments as a function of the frequency of trait use in unconstrained person descriptions (see Table S1)

	Trustworthiness judgments		Dominance judgments	
Number of variables in the model	First PC	Second PC	First PC	Second PC
5	0.90**	0.05	−0.19	0.53**
6	0.91**	0.06	−0.32*	0.77**
7	0.92**	0.01	−0.27*	0.77**
8	0.91**	-.01	−0.24*	0.78**
9	0.90**	-.14	−0.15	0.88**
10	0.91**	-.09	−0.21	0.87**
11	0.92**	-.10	−0.20	0.87**

In the first model, only the five most frequently used trait judgments were entered. Each subsequent model added the next trait judgment according to its frequency of use

*$P < 0.05$.

**$P < 0.001$.

Table S5. Correlations between trait judgments for natural and computer generated faces

Judgments of 66 natural faces	Trustworthy	Dominant	Mean	Threatening
Attractive	0.79	0.04	−0.39	−0.52
Trustworthy		−0.27	−0.72	−0.78
Dominant			0.68	0.68
Mean				0.78
Judgments of 300 computer generated faces	Trustworthy	Dominant	Mean	Threatening
Attractive	0.61	0.15	−0.21	−0.36
Trustworthy		−0.17	−0.67	−0.65
Dominant			0.64	0.68
Mean				0.80

Table S6. Loadings of trait judgments of 300 computer-generated, emotionally neutral faces on the first two PCs

Trait dimension	First PC	Second PC
Trustworthy	0.91	0.03
Frightening	−0.85	0.24
Likeable	0.82	0.39
Threatening	−0.78	0.53
Mean	−0.73	0.57
Attractive	0.73	0.49
Extroverted	0.70	0.18
Competent	0.65	0.67
Dominant	−0.27	0.89
Variance explained	54.3%	25.7%

The loadings represent the correlations of the trait judgments with the PCs.

Table S7. Proportions (standard deviations) of emotion categorization of faces varying on the trustworthiness and dominance dimensions (study 7)

	Trustworthiness dimension				
	−8 SD	−4 SD	0 SD	4 SD	8 SD
Angry	0.67 (0.26)**†	0.20 (0.16)	0.02 (0.05)	0.01 (0.03)	0.00 (0.00)
Disgusted	0.20 (0.24)	0.09 (0.12)	0.01 (0.04)	0.00 (0.00)	0.01 (0.04)
Sad	0.05 (0.10)	0.05 (0.08)	0.01 (0.06)	0.02 (0.06)	0.02 (0.05)
Fearful	0.02 (0.06)	0.06 (0.10)	0.05 (0.07)	0.01 (0.04)	0.00 (0.00)
Neutral	0.05 (0.13)	0.55 (0.26)**†	0.73 (0.22)**†	0.18 (0.19)	0.01 (0.03)
Surprised	0.01 (0.03)	0.05 (0.07)	0.03 (0.07)	0.07 (0.1)	0.11 (0.14)
Happy	0.00 (0.00)	0.01 (0.03)	0.15 (0.17)	0.71 (0.22)**†	0.85 (0.15)**†
	Dominance dimension				
	−8 SD	−4 SD	0 SD	4 SD	8 SD
Angry	0.01 (0.03)	0.00 (0.00)	0.02 (0.05)	0.06 (0.10)	0.24 (0.22)
Disgusted	0.02 (0.06)	0.01 (0.04)	0.01 (0.04)	0.04 (0.06)	0.18 (0.18)
Sad	0.22 (0.20)	0.08 (0.13)	0.01 (0.06)	0.01 (0.03)	0.04 (0.08)
Fearful	0.34 (0.24)*	0.09 (0.14)	0.05 (0.07)	0.03 (0.07)	0.04 (0.07)
Neutral	0.28 (0.28)	0.67 (0.23)**†	0.73 (0.22)**†	0.51 (0.24)**†	0.20 (0.21)
Surprised	0.14 (0.15)	0.09 (0.10)	0.03 (0.07)	0.09 (0.11)	0.13 (0.17)
Happy	0.00 (0.00)	0.05 (0.10)	0.15 (0.17)	0.27 (0.22)*	0.17 (0.25)

The chance level of categorization is 0.14.

*$P < 0.05$.

** $P < 0.001$.

† Passes the significance threshold Bonferroni adjusted for multiple comparisons.